ISO 16269-4:2010-10 (E)

Statistical interpretation of data - Part 4: Detection and treatment of outliers

Contents

Page

Foreword iv

Introduction v

1 Scope 1

2 Terms and definitions 1

3 Symbols 10

4 Outliers in univariate data 11
4.1 General 11
4.1.1 What is an outlier? 11
4.1.2 What are the causes of outliers? 11
4.1.3 Why should outliers be detected? 11
4.2 Data screening 12
4.3 Tests for outliers 14
4.3.1 General 14
4.3.2 Sample from a normal distribution 14
4.3.3 Sample from an exponential distribution 16
4.3.4 Samples taken from some known non-normal distributions 18
4.3.5 Sample taken from unknown distributions 19
4.3.6 Cochran's test for outlying variance 21
4.4 Graphical test of outliers 22

5 Accommodating outliers in univariate data 23
5.1 Robust data analysis 23
5.2 Robust estimation of location 24
5.2.1 General 24
5.2.2 Trimmed mean 24
5.2.3 Biweight location estimate 25
5.3 Robust estimation of dispersion 25
5.3.1 General 25
5.3.2 Median-median absolute pair-wise deviation 25
5.3.3 Biweight scale estimate 26

6 Outliers in multivariate and regression data 26
6.1 General 26
6.2 Outliers in multivariate data 26
6.3 Outliers in linear regression 28
6.3.1 General 28
6.3.2 Linear regression models 29
6.3.3 Detecting outlying Y observations 31
6.3.4 Identifying outlying X observations 31
6.3.5 Detecting influential observations 32
6.3.6 A robust regression procedure 35

Annex A (informative) Algorithm for the GESD outliers detection procedure 36

Annex B (normative) Critical values of outliers test statistics for exponential samples 37

Annex C (normative) Factor values of the modified box plot 44

Annex D (normative) Values of the correction factors for the robust estimators of the scale parameter 47

Annex E (normative) Critical values of Cochran's test statistic 48

Annex F (informative) A structured guide to detection of outliers in univariate data 51

Bibliography 54